Effect of Semi-occluded Vocal Tract Therapy on the Phonation of Children With Vocal Fold Nodules

NCT05878197

22/05/2024

Study protocol

This study was approved by the Ethics Committee of Ghent University Hospital (registration number: B670201936069) and has been registered on ClinicalTrials.gov. The consolidated standards of reporting trials (CONSORT) non-pharmacologic treatment interventions (NPT) checklist were used to report the trial specifications (Boutron et al., 2017). Written informed consent was obtained from a parent of each participant.

Participants

Children between six and 12 years old (primary school) with VFNs were eligible to participate in this study. Participants were recruited between October 2019 and January 2023 during their consultations at the voice unit of Ghent University Hospital, through flyers in other hospital departments, through social media, or by referral from voice therapists who were informed about this project. All participants were diagnosed with VFNs by an otorhinolarygologist with extensive experience in voice diagnostics (S.C.). This age group was chosen because these children can cooperate with voice therapy and tolerate videolaryngostroboscopy. Exclusion criteria were a history of phonosurgery, nasal or ear diseases, neurological disorders, and parent-reported pubertal voice changes.

Design

A randomized controlled study design was used. Participants were assigned to one of the three study arms (experimental group “SP”, experimental group “RVT”, or the control group (CG) receiving indirect treatment), depending on their time of registration. The order in which the different voice treatments were organized was determined using a random number list. No changes in the methodology were made after trial commencement.

Voice therapy

In each study arm, a short-term, intensive, group therapy program was provided. All participants received 11 hours of voice therapy spread over four consecutive days (three hours on day 1-3 and two hours on day 4). Each group consisted of a minimum of two and a maximum of five children and was guided by one experienced voice therapist (I.M. or A.A.). All therapy sessions were provided at Ghent University Hospital. During the study period, participants were not allowed to attend external voice therapy.

The content and hierarchical structure of the different study arms were kept as similar as possible, with the exception of the direct vocal techniques. In the CG, no direct technique was included. Offered therapy consisted of basic education of the vocal mechanism, normal vocal function, and VFNs, vocal hygiene recommendations, posture advice, and costo-abdominal breathing exercises. Participants in the experimental groups received the same counseling and breathing exercises, combined with the direct vocal technique

SP or RVT. For SP, participants used a stainless-steel drinking straw with a diameter of 5 mm and a length of 23 cm. In each study arm, vocal rest pauses and play activities (identical for the experimental groups) were provided. Since no direct vocal technique was practiced in the CG, more play activities were scheduled in which excessive or loud voice use was being monitored and limited. Home exercises were foreseen in each study arm. Parents were not present during therapy but were informed verbally and in writing about the content of the therapy sessions so they could supervise home exercises.

Voice assessment

A standardized and multidimensional voice assessment, based on the protocol of the European Laryngological Society (Dejonckere et al., 2003) and supplemented with the acoustic voice quality index (AVQI) (Maryn et al., 2010a), was performed one day before the start of voice therapy and immediately after the last voice therapy session. A standardized pretest anamnesis by means of a semi-structured interview (see Appendix H) was used to examine the demographic characteristics, therapy history, vocal complaints, and risk factors. The assessments were performed by eight voice therapists of the Center for Speech and Language Sciences (T.P., I.K., J.D., C.L., C.V.S., C.A., L.B., and E.D.) with experience in voice diagnostics in two acoustically treated rooms at Ghent University Hospital. The assessors were blinded to group allocation and participants were assessed by the same assessor pre- and posttherapy. Voice assessment consisted of the following parts: (a) laryngeal anatomy and function, (b) voice quality and vocal capacities, and c) psychosocial wellbeing.

A. Laryngeal anatomy and function

Flexible videolaryngostroboscopy was performed during pretest to determine whether VFNs were present and to describe the baseline laryngeal anatomy and function in detail. In order to avoid overloading the children, videolaryngostroboscopy was performed only pretherapy. It can be perceived as an invasive examination by (young) children, therefore it was considered unethical to perform it twice over a period of five days. Flexible videolaryngostroboscopy was carried out by an experienced otorhinolaryngologist (S.C., more than 25 years of experience in diagnosing voice disorders) or one of her assistants (C.D.V. and M.D.K.). Participants were examined in seated position with the head upright. They were asked to produce a sustained vowel [i] at habitual pitch and loudness, followed by a high-pitched [i] and a low-to-high glissando using [i]. Videolaryngostroboscopy was sometimes terminated early due to discomfort in the children, but recording the vowel [i] at normal pitch and loudness was the minimum requirement. Videolaryngostroboscopic video samples were evaluated using the Voice-Vibratory Assessment with Laryngeal Imaging (VALI) rating form (Poburka et al., 2017) and a grading scale for pediatric VFNs (Nuss et al., 2012; Shah et al., 2007). The video samples were provided with audio. An otorhinolaryngologist (S.C.) and voice therapist (I.K.) randomly and blindly evaluated all video samples. Both raters had experience with administering the VALI rating form, but

not with the grading scale for pediatric VFNs. A half-hour training session was provided in which each parameter was clarified with the definition, a high-quality graphic, and two video examples. The raters first evaluated the video samples independently, followed by a consensus evaluation.

B. Voice quality and vocal capacities

The Dysphonia severity index (DSI) is a multiparametric index, consisting of a weighted combination of the maximum phonation time (MPT, in s), minimal intensity (Ilow, in dB), maximal frequency (fhigh, in Hz), and jitter (%) (Wuyts et al., 2000). MPT was determined by asking the children to sustain the vowel [a:] as long as possible with habitual pitch and loudness after maximum inspiration. Three measurements were performed in free field while participants were seated. Children received verbal encouragements to produce the longest possible vowel. Time was measured with a chronometer and the longest result was used for further analyses. The Ilow and fhigh were determined using the Voice Range Profile of the Computerized Speech Lab (CSL, model 4500, KayPENTAX, Montvale, NY) and a Shure SM-48 microphone, located at a distance of 15 cm from the mouth and angled at 45 degrees. Participants were successively asked to produce the vowel [a:] with minimal loudness and maximal pitch and they were given multiple attempts to achieve the best result. Each production was demonstrated by the assessors, who also verbally encouraged the children. Jitter % was obtained by the Multi-Dimensional Voice Program of the CSL and the same Shure SM-48 microphone. Participants were asked to count to three (an automatic sequence), followed by a sustained vowel [a:] with habitual pitch and loudness. A midvowel segment of three seconds registered with a sampling rate of 50 kHz was used for the analysis. The DSI is calculated using the following formula: 0.13 × MFT + 0.0053 × fhigh - 0.26 × Ilow - 1.18 × jitter + 12.4 (Wuyts et al., 2000). The index value usually ranges between -5 and +5. A lower index value indicates poorer voice quality or, more generally, poorer vocal functioning (Awan & Ensslen, 2010). The threshold score that distinguishes a normophonic from a dysphonic voice is +1.6 in adults (Raes et al., 2002). This threshold score was used since norm values for Dutch children are currently missing.

The Acoustic Voice Quality Index (AVQI) is an objective, multiparameter approach to quantify dysphonia severity on the basis of a sustained vowel [a:] and continuous speech (Maryn et al., 2010a). The index consists of a weighted combination of six time-domain parameters [i.e., shimmer local (SL), shimmer local decibels (SLdB), and harmonics-to-noise ratio (HNR)), frequency-domain parameters (i.e., general slope of the spectrum (slope) and tilt of the regression line through the spectrum (tilt)), and quefrency-domain parameters (i.e., smoothed cepstral peak prominence (CPPs)]. The formula of the AVQI is 2.571 (3.295 − 0.111 CPPs − 0.073 HNR − 0.213 SL + 2.789 SLdB − 0.032 slope + 0.077 tilt). The index ranges from 0 to 10 and a higher index indicates poorer voice quality. The threshold score that distinguishes a normophonic voice from a dysphonic voice in Dutch

adults is 2.95 (Maryn et al., 2010a). An Australian study suggested a threshold score of 3.46 in the pediatric population, but norm values for Dutch children are missing (Reynolds et al., 2012). In order to calculate the AVQI, the participants were asked to produce a sustained vowel [a:] at normal pitch and loudness and to read the Dutch phonetically balanced text "Papa and Marloes" (Van de Weijer & Slis, 1991). A Samson C01U Pro microphone, located at a distance of 15 cm from the mouth, and the software program PRAAT were used (Boersma & Weenink). A midvowel segment of 3 seconds of the vowel [a:] and the first two sentences of the text were combined in an automated script to calculate the AVQI. It is known that a minimum signal-to-noise ratio (SNR) of 30 dB is required to obtain acceptable perturbation measurements (Deliyski et al., 2005). For this reason, all participants with a pre- or posttest sample with an SNR less than 30 dB were removed from the analyses of the AVQI.

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) was used for the auditory-perceptual evaluation of the children's voices (Kempster et al., 2009). This tool is designed to describe the severity of dysphonia by rating six parameters (overall severity, roughness, breathiness, strain, pitch, and loudness) on a visual-analog scale from 0 to 100, where 0 is considered within normal limits. The tool was supplemented by two additional visual-analog scales to measure asthenia and instability (Meerschman et al., 2022). Two experienced voice therapists (T.P. and E.D.) independently and blindly rated all the pseudonymized samples in a random order.
Afterwards, they relistened to the samples together to compare their individual evaluations and to reach a consensus. The samples consisted of the sustained vowel [a:] and the full text "Papa and Marloes," which were also used to determine the AVQI. The raters were allowed to listen multiple times to the samples. Training was not provided because both raters had experience with administering the CAPE-V. A random selection of 15% of the samples was doubled to calculate the intra-rater reliability. For this study, the CAPE-V overall severity was selected as a primary outcome parameter. The analysis of the other CAPE-V parameters can be found in Appendix I.

C. Psychosocial wellbeing

Pediatric Voice Handicap Index (pVHI). A parent completed the Dutch version of the pVHI, which is a parent-proxy tool to gain insight into the physical, functional and emotional impact of the voice disorder on their child (Veder et al., 2017). The pVHI consists of 23 items which are evaluated using a five-point Likert scale (0: never, 1: almost never, 2: sometimes; 3: almost always; 4: always). Total pVHI scores range from 0 to 92 with a higher score corresponding with a greater psychosocial impact. The same parent was asked to complete the questionnaire pre- and posttherapy.

Statistical analysis plan

The statistical analysis was performed using IBM SPSS Statistics 28 software (SPSS, Inc. Chicago, IL). The significance level was set at $\alpha = 0.05$. One-way ANOVAs were used to check for statistically significant differences in continuous parameters (e.g., age) between the three study arms. Fisher's exact tests were used for the comparison of categorical parameters (e.g., sex) between the three study arms. Pairwise comparisons with Bonferroni corrections ($\alpha < 0.016$) were performed when a significant difference was found between the groups. The inter- and intra-rater reliability of the CAPE-V overall severity has been examined with two-way mixed ICCs and their 95% confidence intervals, type consistency, single-rating, and interpreted following the classification of Koo and Li (2016) ((ICC > 0.9 'excellent'; 0.9 ≥ ICC > 0.75 'good'; 0.75 ≥ ICC > 0.5 'moderate'; ICC ≤ 0.5 'poor'). The effects of voice therapy were evaluated using two complementary approaches.

Firstly, a more traditional group-level analysis was conducted. Linear mixed models were used to compare groups over time on the primary outcome parameters, using the restricted maximum likelihood estimation and unstructured covariance type. Time, group, and time-by-group interaction were specified as fixed factors. A random intercept for participants was included and within-group effects of time were determined using pairwise comparisons. Unstandardized effect sizes were presented in the form of the estimated mean differences and 95% confidence intervals for the outcome variables. For the analysis of the AVQI, only participants in whom the SNR of the pretest and posttest voice samples was greater than or equal to 30 dB were selected (Deliyski et al., 2006).

Secondly, an individual-level analysis was conducted for a thorough clinical interpretation of the study results, which is not possible based on group-level average intervention effects (Sand, 2022). This individual analysis examined for each participant whether outcome parameters changed to a relevant degree or remained unchanged. Previous literature was considered to determine the magnitude of this relevant degree of change. For the DSI and AVQI, the relevant degree of change within one patient was set at 2.49 (Hakkesteegt et al., 2008) and 0.54 (Barsties & Maryn, 2013), respectively. For the pVHI and CAPE-V overall severity, the relevant degrees of change were not yet determined. Therefore, it was opted to exclude these outcome parameters from the individual analysis. The results show for DSI and AVQI what percentage of participants improved to a relevant degree, remained unchanged or deteriorated to a relevant degree. Wilson 95% confidence intervals for proportions were calculated.